

### **REMARKS**

In response to the Office Action mailed on September 21, 2009, the Assignee (Nuance Communications, Inc.) respectfully requests reconsideration in view of the foregoing amendments and the following remarks. To further prosecution of this application, each of the rejections set forth in the Office Action has been carefully considered and is addressed below. Claim 16 has been amended to replace a means-plus-function recitation with structural language. Additionally, independent claims 17, 21 and 24 have been amended to make explicit some aspects that are believed to have been implicit in the claims. No new matter has been added. The application is believed to be in condition for allowance.

The Assignee respectfully submits that the amendments to the claims do not require an additional search and/or consideration, such that entry and consideration of the amendments is respectfully requested.

#### **Claim Rejections – 35 U.S.C. §112**

Claims 1-27 stand rejected under 35 U.S.C. §112, first paragraph, as purportedly failing to comply with the written description requirement on the basis that the claims purportedly contain subject matter not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventors, at the time the application was filed, had possession of the claimed invention. (Office Action, page 2). More particularly, it is contended that the features “a speech recognition model of phoneme models” and “phoneme model pairs” are not necessarily known in the art nor appear to be found in the specification. (Office Action, page 2). The Assignee respectfully disagrees.

As one of ordinary skill in the art would have understood when the application was filed, a speech recognition model includes phoneme models. One embodiment of the invention is directed to a technique for generating recognition models. As illustrated in FIG. 1, female training data 104 and male training data 106 is received that contains thousands of recorded phonemes spoken by male and female speakers, where each item of training data is identified by its phoneme(s) and whether it comes from a male speaker or a female speaker. (Page 3, line 27 to page 4, line 3). The specification describes *the phonemes* in the female and male training data as being *modeled* by

quantifying various features from the data, including the data's signal frequencies, intensities, and other characteristics. (Page 4, lines 6-9). As one of ordinary skill would have appreciated, modeling the phonemes results in phoneme models.

The specification also explains that each female model 110 and male model 112 is compared for each phoneme to determine if the gender separation is insignificant. (Page 4, lines 16-18). As one of ordinary skill in the art would have understood, comparing each female model with each male model for each phoneme involves comparing pairs of corresponding phoneme models.

In view of the foregoing, one of ordinary skill in the art would have understood that the inventors, at the time the application was filed, had possession of the claimed invention. Accordingly, the rejection of claims 1-27 under §112 should be withdrawn.

#### Claim Rejections - 35 U.S.C. §103

Each of independent claims 1, 6, 11 and 16 stands rejected under 35 U.S.C. §103(a) as purportedly being unpatentable over Neti (U.S. Patent No. 5,953,701) in view of Yang (U.S. Patent Publication No. 2001/0010039). Each of dependent claims 2-5, 7-10 and 12-15 has been rejected as purportedly being unpatentable over the combination of Neti and Yang, and for some claims, in further view of Kanevsky (U.S. Patent No. 6,529,902).

Each of independent claims 17, 21 and 24 stands rejected under 35 U.S.C. §103(a) as purportedly being unpatentable over Neti in view of Wark (U.S. Patent Publication No. 2003/0231775) and in further view of Yang. Each of dependent claims 18-20, 22-23 and 25-27 has been rejected as purportedly being unpatentable over the combination of Neti, Wark, and Yang.

These rejections are respectfully traversed.

#### A. Overview of Embodiments of the Invention

Speech recognition is the process by which computers analyze sounds and attempt to characterize them as particular letters, words, or phrases. Generally, a speech recognition system is "trained" with many phoneme examples. (Page 1, lines 8-11). A speech recognition system examines various features from each phoneme example by mathematically modeling its sounds on a multidimensional landscape using multiple Gaussian distributions. (Page 1, lines 18-20). Once

acoustic models of phonemes are created, input speech to be recognized is sliced into small samples of sound that are each converted into a multidimensional feature vector by analyzing the same features as previously used to examine the phonemes. (Page 1, lines 21-24). Speech recognition is then performed by statistically matching the feature vector with the closest phoneme model, wherein the accuracy, or word error rate, of a speech recognition system is dependent on how well the acoustic models of phonemes represent the sound samples input by the system. (Page 1, lines 24-29).

Gender specific models, i.e., separate female and male acoustic models of phonemes, are known to yield improved recognition accuracy over gender independent models. (Page 1, line 30 to page 2, line 1). The conventional use of such models is to build one system with just female models and one system with just male models, wherein samples are decoded using both systems in a two-pass approach. (Page 2, lines 1-4). While such gender specific systems provide better speech recognition results, the specification indicates that they generally require too much computing power and resources to be practical in many real-world applications. (Page 2, lines 4-6).

Embodiments of the invention are directed to addressing such limitations of conventional speech recognition systems by generating efficient gender dependent models and integrating such models with an efficient class detection scheme. (Page 2, lines 8-11). Embodiments of the invention determine which models contain class independent information and create class independent models in place of such models. (Page 2, lines 11-13). Embodiments of the invention teach a highly accurate class detection scheme to detect class at a computational cost that is negligible. (Page 2, lines 13-15).

As discussed above, one embodiment of the invention is directed to a technique for generating recognition models. As illustrated in FIG. 1, female training data 104 and male training data 106 is received that contains thousands of recorded phonemes spoken by male and female speakers, where each training data is identified by its phoneme and whether it comes from a male speaker or a female speaker. (Page 3, line 27 to page 4, line 3). The phonemes in the female and male training data are modeled by quantifying various features from the data, including the data's signal frequencies, intensities, and other characteristics. (Page 4, lines 6-9). Female models 110 are

created based solely on the female training data 104 and male models 112 are created based solely on the male training data 106. (Page 4, lines 13-15).

Each female model 110 and male model 112 is compared for each phoneme to determine if the gender separation is insignificant. (Page 4, lines 16-18). For female and male phonemes that are insignificantly different from each other, their female and male training data 104 and 106 are combined and gender independent (GI) models 114 are created. (Page 4, lines 28-31). Additionally, the separate female models 110 and male models 112 that are determined to have insignificant differences from one another are removed. (Page 4, line 31 to page 5, line 1). This results in female models 110 derived from female training data 104, male models 112 derived from male training data 106, and gender independent models 114 derived from both the female and male training data 104 and 106, wherein the female models 110 and male models 112 are significantly different from each other. (Page 5, lines 1-6).

The model creation system and technique beneficially reduces the number of acoustic models of phonemes needed to be stored and searched during speech recognition. (Page 5, lines 7-9). Furthermore, a speech recognition system using the female, male and gender independent models created using this technique requires less computing power, uses less system resources, and is more practical to implement with minimal loss in recognition accuracy. (Page 5, lines 9-13).

FIG. 2 illustrates one process for generating Gaussian Mixture Models (GMMs) according to one embodiment of the invention. The process begins by creating gender independent models ( $GMM_{GI}$ ) from both female and male training data during a training operation 202. (Page 5, lines 24-26). Additionally, female models ( $GMM_f$ ) are created and trained from just the female training data, and male models ( $GMM_m$ ) are created and trained using just the male training data, during training operations 204 and 206. (Page 6, line 27 to page 7, line 1). The various models can be created with training operations that may be performed in sequence or in parallel. After creating and training the models, the female models are compared with the male models to determine if their differences are insignificant, for example, by measuring the differences using the Kullback Leibler divergence. (Page 7, lines 5-10).

For those phonemes which are determined to carry insignificant gender information, the gender independent model for the phoneme is added to a final system model, rather than two gender

dependent models. (Page 9, lines 8-12). For those phonemes that carry gender information determined to be significant, separate female models and male models for phonemes with significant gender information are added to the final system model. (Page 9, lines 13-17). The process continues until examination of all the phoneme models is completed.

In another embodiment, a process for speech recognition that employs the generated female models, male models and gender independent models is illustrated in FIG. 4. More particularly, this aspect of the invention involves a method for recognizing data from a data stream originating from one of a plurality of data classes that includes the female, male and gender independent models described above.

It should be appreciated that the foregoing discussion of embodiments of the invention is provided merely to assist the Examiner in appreciating various aspects of the present invention. However, not all of the description provided above necessarily applies to each of the independent claims pending in the application. Therefore, the Examiner is requested to not rely upon the foregoing summary in interpreting any of the claims or in determining whether they patentably distinguish over the prior art of record, but rather is requested to rely only upon the language of the claims themselves and the arguments specifically related thereto provided below.

B. Each of the Independent Claims Distinguishes Over the Applied References

*Independent claims 1, 6, 11 and 16*

Without acceding to the propriety of the purported combination of Neti and Yang, claims 1, 6, 11 and 16 nevertheless distinguish over these references.

Independent claim 1 is directed to a computer readable medium encoded with instructions that perform a method including determining a difference in model information between pairs of corresponding phoneme models of a female speech recognition model and a male speech recognition model; and creating a gender-independent speech recognition model that includes a gender-independent phoneme model based on a pair of corresponding phoneme models of the female speech recognition model and the male speech recognition model when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant.

Neti relates to gender-dependent models for continuous speech recognition. (Col. 1, lines 9-11). Neti explains that gender dependent speech recognition systems were usually created by splitting or fragmenting training data into each gender and building separate acoustic models for each gender. (Col. 1, lines 14-17). Neti indicates that fragmenting assumes that every state of a sub-phonetic model is uniformly dependent on gender. (Col. 1, lines 17-18). Neti identifies several disadvantages of such gender dependent speech recognition systems, including fragmenting training data when unnecessary, and the need to store a complete acoustic model for each gender. (Col. 1, lines 19-24).

Neti discloses a method of gender dependent speech recognition that includes identifying phone state models common to both genders, identifying gender specific phone state models, identifying a gender of a speaker and recognizing acoustic data from the speaker. (Col. 1, lines 52-56). Neti relates to a method for identifying speech that involves choosing between genders based on gender-dependent sub-phonetic units. (Col. 3, lines 38-40). The method includes a gender question in addition to phone context questions in context decision trees. (Col. 3, lines 44-46). Neti indicates that phone-specific gender dependent acoustic models may be employed using the decision trees. (Col. 3, lines 46-47).

As shown in Fig. 1 of Neti, training data is separated into male and female subsets to create gender-tagged training data, which is then aligned to obtain phonetic context and gender-tag. (Col. 4, lines 10-17). At each node, a gender question and a phonetic-context question is asked by the system, and the question with the best value for the evaluation function is determined. (Col. 4, lines 18-22). If it is determined that the gender question is better, the data is split into two child nodes, and the process of asking the gender and phonetic-context questions and determining the best value for the evaluation function is repeated for each child node. (Col. 4, lines 24-29). Otherwise, the data is split according to the best phonetic-context question. (Col. 4, lines 29-30).

In the Office Action, it is contended that Neti (col. 5, lines 9-21) discloses “determining a difference in model information between pairs of corresponding phoneme models of the female speech recognition model and the male speech recognition model”, and “creating a gender-independent speech recognition model that includes a gender-independent phoneme model based on

a pair of corresponding phoneme models of the female speech recognition model and the male speech recognition model.” (Office Action, page 4). The Assignee respectfully disagrees.

When considered in its entirety, one of ordinary skill in the art would understand that Neti is explaining the use of decision trees, created by asking a question at a node and splitting the result into leaves, to determine the level of gender dependence for constructing a gender-dependent model. Neti does not expressly or inherently teach or suggest determining a difference in model information between pairs of corresponding phoneme models of female and male recognition models, and then creating a gender-independent speech recognition model that includes a gender-independent phoneme model based on a pair of corresponding phoneme models of the female and male recognition models.

In the Office Action, it is further contended that Neti (col. 1, lines 33-47) discloses creating the gender-independent speech recognition model “when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant.” (Office Action, page 4). The Assignee respectfully disagrees.

As one of ordinary skill in the art would appreciate, Neti is describing how acoustic models were built at the time of the Neti application, and further recognizing a need for modeling gender differences that were not sufficiently modeled by context-dependent variations. Neti does not expressly or inherently teach or suggest creating the gender-independent speech recognition model when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant.

Yang fails to cure the deficiencies of Neti. More particularly, Yang is directed to Mandarin Chinese speech recognition by using initial/final phoneme similarity vector. Like Neti, Yang also fails to teach, suggest or even recognize the creation of an independent speech recognition model that includes an independent phoneme model based on a pair of corresponding phoneme models of first and second speech recognition models when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant.

In view of the foregoing, independent claim 1 patentably distinguishes over Neti and Yang, taken either alone or together, which fail to teach or suggest each limitation of the claims.

Accordingly, the rejection of independent claim 1 under §103 as being obvious in view of Neti and Yang should be withdrawn.

Independent claims 6 is directed to a system for generating a speech recognition models that comprises a processing module configured to create an independent speech recognition model that includes an independent phoneme model based on a pair of corresponding phoneme models of a first speech recognition model and a second speech recognition model when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant.

As discussed above, Neti is directed to constructing gender dependent models that involves asking a gender question and a phonetic-context question, and determining the question with the best value for the evaluation function. Neti does not teach or suggest a processing module configured to create an independent speech recognition model that includes an independent phoneme model based on a pair of corresponding phoneme models of a first speech recognition model and a second speech recognition model when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant. Yang fails to cure the deficiencies of Neti.

In view of the foregoing, independent claim 6 patentably distinguishes over Neti and Yang, taken either alone or together, which fail to teach or suggest each limitation of the claims. Accordingly, the rejection of independent claim 6 under §103 as being obvious in view of Neti and Yang should be withdrawn.

Independent claim 11 is directed to a computer program product embodied in computer memory comprising computer readable program codes coupled to the computer memory for generating speech recognition models. The computer readable program codes are configured to cause the program to create an independent speech recognition model that includes an independent phoneme model based on a pair of corresponding phoneme models of a first speech recognition model and a second speech recognition model when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant.

As discussed above, Neti is directed to constructing gender dependent models that involves asking a gender question and a phonetic-context question, and determining the question with the



best value for the evaluation function. Neti does not teach or suggest computer readable program codes that are configured to cause a program to create an independent speech recognition model that includes an independent phoneme model based on a pair of corresponding phoneme models of a first speech recognition model and a second speech recognition model when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant. Yang fails to cure the deficiencies of Neti.

In view of the foregoing, independent claim 11 patentably distinguishes over Neti and Yang, taken either alone or together, which fail to teach or suggest each limitation of the claims. Accordingly, the rejection of independent claim 11 under §103 as being obvious in view of Neti and Yang should be withdrawn.

Independent claim 16 is directed to a system for generating speech recognition models. The system comprises a computer processor programmed to create an independent speech recognition model that includes an independent phoneme model based on a pair of corresponding phoneme models of a first speech recognition model and a second speech recognition model when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant.

As discussed above, Neti is directed to constructing gender dependent models that involves asking a gender question and a phonetic-context question, and determining the question with the best value for the evaluation function. Neti does not teach or suggest a system for generating speech recognition models comprising a computer processor programmed to create an independent speech recognition model that includes an independent phoneme model based on a pair of corresponding phoneme models of a first speech recognition model and a second speech recognition model when the difference in model information between the phoneme models of the pair of corresponding phoneme models is insignificant. Yang fails to cure the deficiencies of Neti.

In view of the foregoing, independent claim 16 patentably distinguishes over Neti and Yang, taken either alone or together, which fail to teach or suggest each limitation of the claims. Accordingly, the rejection of independent claim 16 under §103 as being obvious in view of Neti and Yang should be withdrawn.

*Independent claims 17, 21 and 24*

As amended, independent claim 17 recites a gender-independent speech recognition model that includes independent phoneme models based on pairs of corresponding recorded phonemes originating from the plurality of female speakers and the plurality of male speakers determined to have insignificant differences in model information between the recorded phonemes of the pair of corresponding recorded phonemes. Claim 17 also recites that each of the female speech recognition model and the male speech recognition model lacks the phoneme models of the gender-independent speech recognition model based on pairs of corresponding recorded phonemes originating from the plurality of female speakers and the plurality of male speakers determined to have insignificant differences in model information between the recorded phonemes of pairs of corresponding recorded phonemes.

Independent claims 21 and 24 each recites a third speech recognition model that includes phoneme models based on pairs of corresponding recorded phonemes originating from both the first and second set of speakers determined to have insignificant differences in model information between the recorded phonemes of the pair of corresponding recorded phonemes. Each claim also recites that each of the first speech recognition model and the second speech recognition model lacks the phoneme models of the third speech recognition model based on pairs of corresponding recorded phonemes originating from both the first and second set of speakers determined to have insignificant differences in model information between the recorded phonemes of the pairs of corresponding recorded phonemes.

Without acceding to either the characterization of Wark set forth in the Office Action or the propriety of the purported combination of references, Neti and Yang do not teach or suggest one or more speech recognition models as recited in independent claim 17, 21 and 24. More particularly, the references fail to teach or suggest data classes that include first (female) and second (male) speech recognition models based on recorded phonemes originating from a first set of speakers (female) and a second set of speakers (male), and a third speech recognition model (gender-independent) that includes phoneme models based on pairs of corresponding recorded phonemes originating from first and second sets of speakers determined to have insignificant differences in model information between the recorded phonemes of the pair of corresponding recorded

phonemes, and wherein each of the first and second recognition models lacks the phoneme models of the third speech recognition model based on pairs of corresponding recorded phonemes originating from the first and second sets of speakers determined to have insignificant differences in model information between the recorded phonemes of pairs of corresponding recorded phonemes.

As discussed above, Neti is directed to constructing gender dependent models that involves asking a gender question and a phonetic-context question, and determining the question with the best value for the evaluation function. Neti does not disclose a gender-independent recognition model that is based on pairs of corresponding recorded phonemes originating from the first and second sets of speakers determined to have insignificant differences in model information between the recorded phonemes of pairs of corresponding recorded phonemes. Wark and Yang do not cure at least this deficiency.

In view of the foregoing, independent claims 17, 21 and 24 patentably distinguish over Neti, Wark and Yang, taken either alone or together, which fail to teach or suggest each limitation of the claims. Accordingly, the rejection of independent claim 17, 21 and 24 under §103 as being obvious in view of Neti, Wark and Yang should be withdrawn.

#### *Dependent claims*

Each of the dependent claims depends from one of independent claims 1, 6, 11, 16, 17, 21 and 24 and is patentable for at least the same reasons. Thus, the rejection of each of the dependent claims should similarly be withdrawn.

Since each of the dependent claims depends from a base claim that is believed to be in condition for allowance, the Assignee believes that it is unnecessary at this time to argue the further distinguishing features of the dependent claims. However, the Assignee does not necessarily concur with the interpretation of the dependent claims set forth in the Office Action, nor does the Assignee concede that the prior art alleged to show the features in the dependent claims does so. Therefore, the Assignee reserves the right to specifically address the further patentability of the dependent claims in the future.

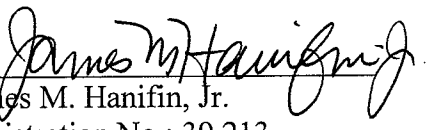
**CONCLUSION**

In view of the foregoing amendments and remarks, this application should now be in condition for allowance. A notice to this effect is respectfully requested. If the Examiner believes, after this amendment, that the application is not in condition for allowance, the Examiner is requested to call the undersigned at the telephone number listed below.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, the Assignee hereby requests any necessary extension of time. If there is a fee occasioned by this response, including an extension fee, the Director is hereby authorized to charge any deficiency or credit any overpayment in the fees filed, asserted to be filed or which should have been filed herewith to our Deposit Account No. 23/2825, under Docket No. N0484.70762US00.

Dated: 11/23/09

Respectfully submitted,  
Nuance Communications, Inc.

By   
James M. Hanifin, Jr.  
Registration No.: 39,213  
Richard F. Giunta  
Registration No.: 36,149  
WOLF, GREENFIELD & SACKS, P.C.  
Federal Reserve Plaza  
600 Atlantic Avenue  
Boston, Massachusetts 02210-2206  
617.646.8000